Experimental Design for Immunologists

Hulin Wu, Ph.D., Dean’s Professor
Department of Biostatistics & Computational Biology
Co-Director: Center for Biodefense Immune Modeling
School of Medicine and Dentistry
University of Rochester
Biostatistics vs. Lab Research
Outline

- Introduction
- Hypothesis: Formulation and statement of the problem
- Experimental design
- Perform experiments to collect data
- Data management, processing and analysis
- Result interpretation and annotation
- Standard experimental design and complicated experimental design: Examples
- Experiment design for systems biology and immune modeling
Biomedical Investigation Process

- **Hypothesis:** Formulation and statement of the problem
- **Design of experiment**
- **Performance of experiment and data collection**
- **Data management, processing and analysis**
- **Result interpretation and annotation**
I. Generate hypotheses: formulation and statement of the problem

- Based on some hypotheses, what scientific questions do you want to address from the experiment?
- Formulate the question in a concrete way instead of a ‘general’ term
- Consider the key endpoints: response (or dependent) variables
- Consider the key controllable variables, factors and experimental conditions: independent variables or covariates
I. Generate hypotheses: formulation and statement of the problem

- Based on some hypotheses, what scientific questions do you want to address from the experiment?

  Example: study whether an adjuvant can enhance H5N1 vaccine response
I. Generate hypotheses: formulation and statement of the problem

- Formulate the question and statement of the problem
  
  - Example 1: Does an H5N1 vaccine with an adjuvant work better?
  
  - Example 2: Is there a difference in cellular biomarker responses measured by flow cytometry, ELISPOT, HAI titer of anti-H5N1 in primary murine H5N1 vaccination with and without an adjuvant
I. Generate hypotheses: formulation and statement of the problem

- Spell out the problem to be well understood and point the way how to conduct the experiments
- Key endpoints: response or dependent variables
  - Cellular biomarker responses measured by flow cytometry, ELISPOT, HAI titer of anti-H5N1
  - Data: continuous, binary, categorical, ordinal?
I. Generate hypotheses: formulation and statement of the problem

- Key controllable variables, factors and experimental conditions: independent variables or covariates
  - Primary murine H5N1 vaccination with and without an adjuvant
  - Any other factors and experimental conditions need to be considered?
  - For example: dose of the vaccine, a boost dose at later day, the sampling time post vaccination, sampling organs (lung, spleen, blood, draining lymph nodes…) etc.
II. Design of experiments: Three Types of Variables or Factors

- Independent variables or factors
  - Manipulable at levels of interest
  - Qualitative/discrete or quantitative/continuous
  - Having an impact upon the dependent/response variables (endpoints)

- Dependent or response variables: study endpoints
  - Easily and efficiently measurable
  - Related to the scientific questions

- Nuisance or supplementary variables: no interest
  - May not be measurable
  - May be unknown, but potentially confounding
  - Not easy to control
II. Design of experiments: Max-Min-Con Principle

- **Max**imize the variance of the independent variables: a large variation of controllable independent variables → a larger effect
- **Min**imize error variance: dependent variables
  1. Standardized procedures
  2. Reliable individual measure
  3. Large samples
  4. Using an homogeneous sample to reduce the error variance due to individual difference variables.
- **Con**trol extraneous variance: nuisance or supplementary variables
  1. Eliminate the nuisance variable (fix as constant).
  2. Use randomization to control the effect of nuisance variables.
  3. Build the nuisance variable into the research design.
II. Design of experiments: Manipulated independent factors

- Set at levels of interest to maximize the effects of interest
  - A large range or more levels of controllable independent variables

- Examples: H5N1 vaccine study
  - Dose of the vaccine: a large range or more levels, but need to consider tolerability
II. Design of experiments: Dependent/Response Variables

- Minimize error variance
- Examples: H5N1 vaccine study
  - Use the same equipment (flow cytometer), the same batch of reagents, experimental conditions etc.
  - Use the same type of mice
  - Standardize the assay and the protocol
  - Use automatic and objective operation: e.g. flow cytometry data gating
  - Use replicates
II. Design of experiments: Examples

- Control or fixed nuisance variables throughout the experiment: the study of H5N1 vaccine with adjuvant
  - Use the same age and weight of mice or
  - Consider the age and weight of mice as control factors in the design
II. Design of experiments: Randomization

- Average out the effects of variables or factors that cannot be controlled: nuisance variables
  
  - Many unknown factors or factors cannot be controlled
  - Randomization
  - Blinded study: single-blind and double-blind
  - Examples: H5N1 vaccine study
    - Randomize mice for the two arms:
    - H5N1 vaccine without adjuvant vs. H5N1 with adjuvant
    - Randomize the samples into different labs or technicians
    - Randomize the order of measurements
II. Design of experiments: Sample size and observations:

- How many subjects and how many data points or replicates per subject? The time or location of measurements?
- Need the information:
  - Effect size to be detected
  - Variation of the endpoint
  - Power: 80%
  - Significance level: 0.05
  - Software for power justification or sample size calculation
II. Design of experiments: Sample size and observations:

- If no prior information, take as large a sample as possible
  - Experimental feasibility
  - Experimental efficiency and cost-effectiveness
- Optimal design: the location and time of the observations
II. Design of experiments: Statistical Model

- Formulate the scientific question into a statistical framework or statistical model based on above design conditions
  - Hypothesis testing
  - ANOVA
  - Correlation
  - Regression
  - .......
III. Perform experiments to collect data

- Perform randomization
- Keep consistency
- Keep the Max-Min-Con principle
IV. Data management, processing and analysis

- Data management plan
  - Centralize the data into a relational database with quality control and management

- Data processing and analysis plan

- Perform data processing and analysis
  - Based on the experimental design and assumptions

- Data analysis plan: important

- Data analysis report

- Draft a manuscript for publication
IV. Data Analysis

- Descriptive statistics and graphical displays of the data
- Standard data analysis and statistical inference
  - Hypothesis testing
  - ANOVA
  - Correlation
  - Regression
  - ……..
- Advanced modeling and data analysis
V. Result interpretation and annotation

- Collaboration between biomedical scientists and statisticians/bioinformaticians to interpret the results and annotate the data
- Draft manuscripts for publication
- Iterate Step I-V
Standard Experimental Design: Compare Two or More Groups

- Single-factor design: One response variable
- Design of experiment: simple, but randomization to control nuisance variables
- Statistical model: One-Way ANOVA
- Randomized block design: Two-Way ANOVA
- Examples:
  - H5N1 vaccine study: control, dose 1, 2 & 3
  - Randomized block: 4 labs
One-Way ANOVA and Randomized Block Design

- **One-Way ANOVA**: One lab for 4 doses

<table>
<thead>
<tr>
<th></th>
<th>D1</th>
<th>D2</th>
<th>D3</th>
<th>D4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>24</td>
<td>32</td>
<td>43</td>
<td>49</td>
</tr>
<tr>
<td>2</td>
<td>11</td>
<td>21</td>
<td>39</td>
<td>58</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>39</td>
<td>48</td>
<td>42</td>
</tr>
<tr>
<td>4</td>
<td>17</td>
<td>33</td>
<td>51</td>
<td>63</td>
</tr>
</tbody>
</table>

- **Randomized block design**: 4 labs & 4 doses

<table>
<thead>
<tr>
<th></th>
<th>L1</th>
<th>D1</th>
<th>D2</th>
<th>D3</th>
<th>D4</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1</td>
<td>D1</td>
<td>D2</td>
<td>D3</td>
<td>D4</td>
<td></td>
</tr>
<tr>
<td>L2</td>
<td>D1</td>
<td>D2</td>
<td>D3</td>
<td>D4</td>
<td></td>
</tr>
<tr>
<td>L3</td>
<td>D1</td>
<td>D2</td>
<td>D3</td>
<td>D4</td>
<td></td>
</tr>
<tr>
<td>L4</td>
<td>D1</td>
<td>D2</td>
<td>D3</td>
<td>D4</td>
<td></td>
</tr>
</tbody>
</table>
Standard Experimental Design: Multiple Factor Design

- Factorial design: Two-Way ANOVA
- Latin square design: More factors
- Examples: 4x4 Latin Square

\[
\begin{array}{cccc}
1 & 2 & 3 & 4 \\
1 & A & B & D & C \\
2 & D & C & A & B \\
3 & B & D & C & A \\
4 & C & A & B & D \\
\end{array}
\]

A=Dose 0, B=Dose 1, C=Dose 2, D=Dose 3
Standard Experimental Design: Factorial Design

- Full factorial design: Consider 7 factors with two levels for each factor \(2^7 = 128\) runs

- Fractional factorial design
  - Reduce the number of runs or redundancy
  - Ignore higher-order interactions
Complicated Experimental Design

- Regression and generalized regression
- Time-to-event or survival endpoint
- Longitudinal studies
- High-throughput bioinformatics studies
- Immune modeling: Differential equation models
Complicated Experimental Design

- All basic (Max-Min-Con) principles regarding experimental design apply
- Sample size and observation location/time determination: more complicated
- Computer simulations: necessary to explore feasibility and power calculations
Design for Immune Modeling and Systems Biology Experiments

- Use systems biology approaches to design immune modeling experiments:
  - Study the interactions between the components of biological systems
  - A new perspective: integration rather than reduction
  - Experiments: measuring multiple components simultaneously
  - New techniques: a large amount of high-throughput `omics' data
Challenges for systems biology experiments

- Complex high-throughput data at multi-scale (genetic, protein, cellular, organ, and population) levels
- Complex relationships of many factors in a biological system: nonlinear
- Dynamics, instead of static system
- Need support from interdisciplinary quantitative sciences: Rigorous data integration with mathematical models
Design of Systems Biology
Experiments: Example

- H5N1 Vaccine Study
  - Measure cellular level functions: Elispot, Flow cytometry etc.
  - Measure cytokines, chemokines and protein expressions: Elisa, Luminex, proteomics
  - Measure gene expressions: microarray, RNA-Seq
  - Time Course (dynamics): Days 0, 1, 2, … 10,
  - Models: Differential equations
  - Design simulations
Summary

- Do not separate experimental design and data analysis
- Check your model assumptions
- Make your design efficient and cost-effective
- Make your design novel
- Consult and collaborate with statisticians at the early stage of experimental design